AD693991

NUSL Report No. 1032

Numerical Evaluation of Cumulative Probability Distribution Functions Directly from Characteristic Functions

ALBERT H. NUTTALL

Office of the Associate Technical Director for Research

11 August 1969

This document has been approved for public release and sale; its distribution is unlimited.

Navy Underwater Sound Laboratory
New London, Connecticut 06320

ABSTRACT

A method for direct numerical evaluation of the cumulative probability distribution function from the characteristic function in terms of a single integral is presented. No moment evaluations or series expansions are required. Intermediate evaluation of the probability density function is circumvented. The method takes on a special form when the random variables are discrete.

ADMINISTRATIVE INFORMATION

This study was performed under USL Project No. A-041-00-00, Navy Subproject and Task No. ZF 11 112 002.

TABLE OF CONTENTS

LIST OF TABLES

NUMERICAL EVALUATION OF CUMULATIVE PROBABILITY DISTRIBUTION FUNCTIONS DIRECTLY FROM CHARACTERISTIC FUNCTIONS

INTRODUCTION

When several independent random variables are added, the characteristic function of the sum is the product of the characteristic functions of the individual random variables. This rule holds regardless of the distributions of the individual random variables, and whether they are identically distributed or not. Evaluation of the cumulative probability distribution of the sum variable in closed form is often very tedious or impossible to achieve. This is especially so when the number of random variables added is large, but not large enough to employ the Central Limit Theorem with accuracy.

In many signal-detection problems, the characteristic function of the decision variable can be derived in closed form (or evaluated numerically fairly easily). Often, however, neither the probability density function of the decision variable, nor its integral, the cumulative probability distribution function, can be obtained in closed form. Even if they can, they are frequently tedious and time-consuming to evaluate (see, for example, Marcum[1]). In this report, we present a technique for numerically evaluating cumulative probability distribution functions directly from specified characteristic functions in terms of a single integral. Intermediate evaluations of the probability density functions are not necessary, and no moment evaluations or series expansions are required. The technique takes on a special form when the decision variable is discrete.

When the characteristic function of the decision variable (which is compared with a threshold) can be evaluated for both the signal-present and signal-absent cases, the technique can be applied to the problem of obtaining receiver operating characteristics (probability of detection versus probability of false alarm).

ANALYSIS

This section is composed of two subsections. In the first, a general formula for direct evaluation of the cumulative probability distribution function from the characteristic function is derived; in the second, an alternate and more useful form for discrete random variables is presented.

GENERAL DISTRIBUTIONS

Let random variable x have probability density function (PDF) p(x) and characteristic function (CF) $f(\xi)$:

$$f(\xi) = \int dx \exp(i\xi x)\, p(x), \tag{1}$$

and

$$p(x) = \frac{1}{2\pi} \int d\xi \exp(-i\xi x)\, f(\xi). \tag{2}$$

(An integral without limits is over the real axis from $-\infty$ to $+\infty$.)

The cumulative distribution function (CDF) $\Pr(X)$ is defined as the probability that random variable x is less than or equal to X:

$$\Pr(X) = \int_{-\infty}^{X+} dx\, p(x). \tag{3}$$

The upper limit means that an impulse in PDF p(x) at $x = X$ is to be included in full. It will be convenient to define the modified distribution function (MDF):

$$P(X) = \int_{-\infty}^{X} dx\, p(x), \tag{4}$$

where an impulse in p(x) at $x = X$ is only half included. At points of continuity of the CDF, $\Pr(X)$ and $P(X)$ are equal. At a point of discontinuity of the CDF, the MDF $P(X)$ takes on a value halfway between the limit values on either side of the discontinuity.[2] The CDF $\Pr(X)$ can be obtained from the MDF $P(X)$ via

$$\Pr(X) = \lim_{\epsilon \to 0+} P(X + \epsilon). \tag{5}$$

Therefore, we can direct our effort to evaluating either the CDF $\Pr(X)$ or the MDF $P(X)$, depending on which is more convenient.

When Eq. (2) is substituted into Eq. (4), we note that the MDF becomes[3]

$$
\begin{aligned}
P(X) &= \int_{-\infty}^{X} dx \frac{1}{2\pi} \int d\xi \exp(-i\xi x) f(\xi) \\
&= \frac{1}{2\pi} \int d\xi\, f(\xi) \int_{-\infty}^{X} dx \exp(-i\xi x) \\
&= \frac{1}{2\pi} \int d\xi\, f(\xi) \left[\pi\delta(\xi) - \frac{1}{i\xi} \exp(-i\xi X) \right] \\
&= \frac{1}{2} - \frac{1}{i2\pi} \int \frac{d\xi}{\xi} f(\xi) \exp(-i\xi X),
\end{aligned}
\tag{6}
$$

where the last integral is a principal value integral. Since the PDF $p(x)$ is real, the real part of the CF $f(\xi)$ is even, and the imaginary part of the CF $f(\xi)$ is odd; i.e., $f(-\xi) = f^*(\xi)$. This allows Eq. (6) to be manipulated into the forms

$$
\begin{aligned}
P(X) &= \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \frac{d\xi}{\xi} \operatorname{Im}\left\{ f(\xi) \exp(-i\xi X) \right\} \\
&= \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \frac{d\xi}{\xi} \left[\operatorname{Im}\left\{ f(\xi) \right\} \cos(\xi X) - \operatorname{Re}\left\{ f(\xi) \right\} \sin(\xi X) \right].
\end{aligned}
\tag{7}
$$

Convergence of the integrals† at the origin is guaranteed by the fact that

† ξ^ν is integrable at the origin if $\nu > -1$. No moments of the distribution are required to exist.

$$\mathrm{Im}\left\{ f(0) \right\} = 0 \tag{8}$$

and

$$\lim_{\xi \to 0} \frac{\sin(\xi \mathbf{X})}{\xi} = \mathbf{X}. \tag{9}$$

Equation (7) is the general equation allowing numerical evaluation of the MDF $P(\mathbf{X})$ directly from the CF $f(\xi)$. For a discontinuous CDF, in order to minimize inaccuracies in a numerical evaluation of Eq. (7), values of the MDF $P(\mathbf{X})$ at points removed from the discontinuity locations (if known) should be computed. In particular, for a discrete random variable, values of the MDF at points midway between discontinuities should be computed when using Eq. (7).

The integral in Eq. (7) is confined to the real axis. Since

$$\left| f(\xi) \right| \le \int dx\, p(x) = 1 \qquad \text{for } \xi \text{ real}, \tag{10}$$

there are no singular points along the ξ axis. Also some CF's are defined only for ξ real; for example, for

$$p(x) = \frac{1}{\pi} \frac{1}{1+x^2}, \tag{11}$$

$$f(\xi) = \exp(-|\xi|), \qquad \text{real } \xi, \tag{12}$$

but $f(\xi)$ is not defined for complex ξ. Thus, the CF $f(\xi)$ does not have to be analytic at the origin to apply Eq. (7). Nor do any moments of the random variable have to exist.

DISCRETE DISTRIBUTIONS

The expression (7) applies to all MDF's (and CDF's through Eq. (6)); however, it requires an infinite integral for each value of $\mathbf{X}$. Here we shall alleviate this requirement for a special class of random variables. Namely, we consider discrete random variables that can only take on values which are multiples of some fundamental increment Δ. That is, the PDF of interest takes the form

$$p(x) = \sum_k c_k\, \delta(x - k\Delta). \tag{13}$$

(A sum without limits is over the integers from $-\infty$ to $+\infty$.) Then the CF is

$$f(\xi) = \sum_k c_k \exp(ik\Delta\xi), \tag{14}$$

which is periodic with period $2\pi/\Delta$. Therefore, the coefficients $\{c_k\}$ can be determined from the CF $f(\xi)$ by

$$c_k = \frac{\Delta}{2\pi} \int_{2\pi/\Delta} d\xi \exp(-ik\Delta\xi)\, f(\xi), \tag{15}$$

where the integral is over any interval of length $2\pi/\Delta$.

Equation (15) gives the area of any impulse in the PDF $p(x)$ in terms of a finite integral of the CF $f(\xi)$. Since we are interested in the CDF $\Pr(X)$, a sum over $\{c_k\}$ is required. At this point, it is convenient to distinguish two cases: (1) nonnegative discrete random variables and (2) general discrete random variables.

NONNEGATIVE DISCRETE RANDOM VARIABLES

If x is a nonnegative discrete random variable, the CDF is, at integer value M,

$$\Pr(M) = \sum_{k=0}^{M} c_k = \frac{\Delta}{2\pi} \int_{2\pi/\Delta} d\xi\, f(\xi) \sum_{k=0}^{M} \exp(-ik\Delta\xi), \quad M \geq 0, \tag{16}$$

where we have substituted Eq. (15). Now

$$\sum_{k=0}^{M} \exp(-ik\Delta\xi) = \frac{1 - \exp[-i(M+1)\Delta\xi]}{1 - \exp[-i\Delta\xi]}. \tag{17}$$

which must be interpreted as $M + 1$ at $\xi = 0, \pm 2\pi/\Delta, +4\pi/\Delta, \ldots$. Using Eq. (17) and the fact that $f(-\xi) = f^*(\xi)$, we note that Eq. (16) becomes

$$\Pr(M) = \frac{\Delta}{2\pi} \int_{2\pi/\Delta} d\xi \; f(\xi) \exp(-iM\Delta\xi/2) \frac{\sin\left[(M+1)\Delta\xi/2\right]}{\sin\left[\Delta\xi/2\right]}$$

$$= \frac{\Delta}{\pi} \int_0^{\pi/\Delta} d\xi \frac{\sin\left[(M+1)\Delta\xi/2\right]}{\sin\left[\Delta\xi/2\right]} \operatorname{Re}\left\{f(\xi)\exp(-iM\Delta\xi/2)\right\}, \quad M \geq 0, \tag{18}$$

where the interval $(-\pi/\Delta, \pi/\Delta)$ has been selected for integration. The ratio of sines is interpreted as $M + 1$ at the origin $\xi = 0$. Equation (18) is a single finite integral from which the CDF $\Pr(M)$ can be evaluated at any M directly from the CF $f(\xi)$.

A special case of Eq. (18) is

$$\Pr(0) = c_0 = \frac{\Delta}{\pi} \int_0^{\pi/\Delta} d\xi \operatorname{Re}\left\{f(\xi)\right\}. \tag{19}$$

(Actually, c_0 is always given by this formula, even for general discrete random variables, as may be seen from the general formula (Eq. (15)).)

The case of a discrete random variable taking on values in a semi-infinite range (i.e., $(-\infty, N)$ or (N, ∞), where N is finite but can be positive or negative) can be handled in a similar fashion. The key is that a finite sum of exponentials (like Eq. (17)) can be evaluated without requiring a summation.

GENERAL DISCRETE RANDOM VARIABLES

Here we shall consider discrete random variables which can take on values in the range $(-\infty, \infty)$. From Eqs. (7), (4), and (13),

$$P(0) = \frac{1}{2} - \frac{1}{\pi} \int_0^{\infty} \frac{d\xi}{\xi} \operatorname{Im}\left\{f(\xi)\right\} \tag{20}$$

$$= \sum_{k=-\infty}^{-1} c_k + \frac{1}{2} c_0 \tag{21}$$

That is, the value of the MDF $P(\mathbf{X})$ at the origin can be evaluated by a single infinite integral. There does not seem to be any simpler way of obtaining this number, which will be necessary in the development to follow. In some cases, it may be possible to evaluate the particular value $P(0)$ from the integral Eq. (20) in closed form, or expand it in a rapidly convergent series, while $P(\mathbf{X})$ could not be so evaluated generally for $\mathbf{X} \neq 0$. In any event, Eq. (20) will be the only infinite integral necessary to evaluate in order to get the complete CDF for this general discrete case.

The area of the impulse at the origin is given by Eq. (15) as

$$c_0 = \frac{\Delta}{\pi} \int_0^{\pi/\Delta} d\xi \, \mathrm{Re}\left\{f(\xi)\right\}. \tag{22}$$

Now let us define auxiliary functions

$$S_+(M) = \sum_{k=0}^{M} c_k , \quad M \geq 0 , \tag{23}$$

$$S_-(M) = \sum_{k=-M}^{0} c_k , \quad M \geq 0 . \tag{24}$$

By a development similar to Eqs. (16) through (18), we find that these auxiliary functions can be expressed directly in terms of the CF $f(\xi)$ as

$$S_{\pm}(M) = \frac{\Delta}{\pi} \int_0^{\pi/\Delta} d\xi \, \frac{\sin\left[(M+1)\Delta\xi/2\right]}{\sin\left[\Delta\xi/2\right]} \, \mathrm{Re}\left\{f(\xi)\exp(\mp iM\Delta\xi/2)\right\}, \; M \geq 0, \tag{25}$$

where the ratio of sines is interpreted as $M+1$ at the origin $\xi = 0$.

The CDF $\Pr(M)$ then can be evaluated at any M according to

$$\Pr(M) = \left\{ \begin{array}{ll} P(0) - \frac{1}{2}c_0 + S_+(M), & M \geq 0 \\ P(0) + \frac{1}{2}c_0 - S_-(|M+1|), & M < 0 \end{array} \right\} . \tag{26}$$

Here $P(0)$ is given by Eq. (20), c_0 by Eq. (22), and $S_+(M)$ by Eq. (25). The constants $P(0)$ and c_0 need be evaluated once, but Eq. (25) must be evaluated for each M of interest. However, Eq. (25) is a finite integral.

EXAMPLES

We shall consider two examples recently examined by Helstrom[4] for purposes of comparison.

Example 1 - Exponential Distribution

$$p(x) = \begin{cases} \exp(-x), & x \geq 0 \\ 0, & x < 0 \end{cases}, \tag{27}$$

$$Pr(\mathbf{X}) = P(\mathbf{X}) = \begin{cases} 1 - \exp(-\mathbf{X}), & \mathbf{X} \geq 0 \\ 0, & \mathbf{X} < 0 \end{cases}, \tag{28}$$

$$f(\xi) = (1-i\xi)^{-1} . \tag{29}$$

The exact CDF is given in Eq. (28). Approximate values for the CDF are obtained by substituting Eq. (29) into Eq. (7) and approximating the infinite integral by a finite sum. Results are indicated in Table 1.

The integral of Eq. (7) was sampled in ξ at values indicated by column four of Table 1 and approximated by the trapezoidal rule for integration. The limit of integration in Eq. (7) was taken to be the value above 60 where the finite sum deviated <u>most</u> from the exact answer. Thus, the finite sum in column three of Table 1 is the worst approximation to the exact answer in column two.

For this example, the largest error occurred at the origin. This happened because the integrand of Eq. (7) oscillates for $\mathbf{X} \neq 0$, thereby converging fairly rapidly, whereas the integrand decreases monotonically only as $(1 + \xi^2)^{-1}$ for $\mathbf{X} = 0$.

Table 1

NUMERICAL COMPUTATION OF EXPONENTIAL DISTRIBUTION

X	Pr(X)	Finite Sum via Eq. (7)	Increment in ξ	Approximate Limit of Integration
-10	0	.00001	.1	60
-2	0	-.00007	.5	60
-1	0	.00008	.5	60
0	0	.00532	.5	60
.2	.18127	.18096	.5	60
1	.63212	.63220	.5	60
2	.86466	.86470	.5	60
10	.9999546	.9999637	.1	60

Example 2 - Poisson Distribution

$$p(x) = \exp(-\lambda) \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \delta(x-k), \tag{30}$$

$$\Pr(M) = \left\{ \begin{array}{r} \exp(-\lambda) \sum_{k=0}^{M} \frac{\lambda^k}{k!}, \quad M \geq 0 \\ 0, \quad M < 0 \end{array} \right\}, \tag{31}$$

$$f(\xi) = \exp\left[\lambda\left\{\exp(i\xi)-1\right\}\right] . \tag{32}$$

The exact CDF is given in Eq. (31). Approximate values for the CDF are obtained by substituting Eq. (32) into Eq. (18), with $\Delta = 1$, and approximating the finite integral by a finite sum. Results are indicated in Table 2.

Table 2

NUMERICAL COMPUTATION OF POISSON DISTRIBUTION

M	Pr(M)	Finite Sum via Eq. (18)	Number of Intervals
0	.00000 03059	.00000 03059	25
1	.00000 48944	.00000 48945	25
6	.00763 18996	.00763 18998	25
14	.46565 37089	.46565 37089	25
16	.66412 32005	.66412 32004	25
20	.91702 90899	.91702 90895	25
29	.99958 15502	.99958 15500	25
30	.99980 26867	.99980 26865	25
40	.99999 99765	.99999 99764	25

The integral of Eq. (18) was divided into 25 equal intervals and approximated by the trapezoidal rule for integration. Columns two and three of Table 2 show that the error in the approximation occurs in the tenth place (and may be due to computer inaccuracies rather than sampling errors). Also, the accuracy holds on the tails of the CDF as well as near the mean.

CONCLUSIONS

The numerical technique suggested for obtaining CDF's directly from CF's has considerable merit. It requires no moment evaluations or series expansions (like Edgeworth or Laguerre) for the distributions. It does not depend upon evaluation of derivatives of CF's, but depends only upon the values of the CF itself. (Evaluation of high-order derivatives can be extremely tedious and time-consuming even if an analytic form for the CF is available.) The accuracy of the suggested technique can be estimated and controlled by decreasing the increment in the integral evaluations or lengthening the interval of integration or both; the change in the approximation is a measure of the error at that point. The method does not require an inordinate number of samples of the CF, at least for the examples considered, and the additional functions requiring evaluation are sines and cosines. Intermediate evaluation of the PDF is entirely circumvented. (Of course, estimates of the PDF are available as differences of the CDF, if desired.)

LIST OF REFERENCES

1. J. I. Marcum, A Statistical Theory of Target Detection by Pulsed Radar: Mathematical Appendix, Rand Corporation Report No. RM-753, 1 July 1948.

2. A. Papoulis, The Fourier Integral and Its Applications, McGraw-Hill Book Co., Inc., New York, 1962, sect. 2.1.

3. Ibid, Eqs. (3-12), (3-6), (3-9), and (2-36) and sect. 2.1.

4. C. W. Helstrom, "Approximate Calculation of Cumulative Probability from a Moment-Generating Function," Proceedings of the IEEE, vol. 57, no. 3, March 1969, pp. 368-369.

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Navy Underwater Sound Laboratory New London, Connecticut 06320	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP

3. REPORT TITLE
NUMERICAL EVALUATION OF CUMULATIVE PROBABILITY DISTRIBUTION FUNCTIONS DIRECTLY FROM CHARACTERISTIC FUNCTIONS

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
Research Report

5. AUTHOR(S) (First name, middle initial, last name)
Albert H. Nuttall

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
11 August 1969	18	4

8a. CONTRACT OR GRANT NO. b. PROJECT NO. A-041-00-00 c. ZF-11-112-002 d.	9a. ORIGINATOR'S REPORT NUMBER(S) 1032 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT
This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Department of the Navy

13. ABSTRACT

A method for direct numerical evaluation of the cumulative probability distribution function from the characteristic function in terms of a single integral is presented. No moment evaluations or series expansions are required. Intermediate evaluation of the probability density function is circumvented. The method takes on a special form when the random variables are discrete.

DD FORM 1 NOV 65 1473 (PAGE 1)
S/N 0102-014-6600

UNCLASSIFIED
Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Discrete Random Variables						
Characteristic Function						
Cumulative Probability Distribution Function						
Numerical Evaluation						
Probabilities of Detection and False Alarm						
Receiver Operating Characteristics						

ABSTRACT CARD LAYOUT

3ND-USNUSL-232

Navy Underwater Sound Laboratory
Report No. 1032
NUMERICAL EVALUATION OF CUMULATIVE PROBABILITY DISTRIBUTION FUNCTIONS DIRECTLY FROM CHARACTERISTIC FUNCTIONS, by Albert H. Nuttall. 11 August 1969. i-iv + 14 p., tables. UNCLASSIFIED

A method for direct numerical evaluation of the cumulative probability distribution function from the characteristic function in terms of a single integral is presented No moment evaluations or series expansions are required Intermediate evaluation of the probability density function is circumvented. The method takes on a special form when the random variables are discrete.

1. Statistical Functions
I. Nuttall, A.H.
II. Title
III. ZF 11 112 002

Navy Underwater Sound Laboratory
Report No. 1032
NUMERICAL EVALUATION OF CUMULATIVE PROBABILITY DISTRIBUTION FUNCTIONS DIRECTLY FROM CHARACTERISTIC FUNCTIONS, by Albert H. Nuttall. 11 August 1969. i-iv + 14 p., tables. UNCLASSIFIED

A method for direct numerical evaluation of the cumulative probability distribution function from the characteristic function in terms of a single integral is presented No moment evaluations or series expansions are required Intermediate evaluation of the probability density function is circumvented. The method takes on a special form when the random variables are discrete.

1. Statistical Functions
I. Nuttall, A.H.
II. Title
III. ZF 11 112 002

Navy Underwater Sound Laboratory
Report No. 1032
NUMERICAL EVALUATION OF CUMULATIVE PROBABILITY DISTRIBUTION FUNCTIONS DIRECTLY FROM CHARACTERISTIC FUNCTIONS, by Albert H. Nuttall. 11 August 1969. i-iv + 14 p., tables. UNCLASSIFIED

A method for direct numerical evaluation of the cumulative probability distribution function from the characteristic function in terms of a single integral is presented No moment evaluations or series expansions are required Intermediate evaluation of the probability density function is circumvented. The method takes on a special form when the random variables are discrete.

1. Statistical Functions
I. Nuttall, A.H.
II. Title
III. ZF 11 112 002

Navy Underwater Sound Laboratory
Report No. 1032
NUMERICAL EVALUATION OF CUMULATIVE PROBABILITY DISTRIBUTION FUNCTIONS DIRECTLY FROM CHARACTERISTIC FUNCTIONS, by Albert H. Nuttall. 11 August 1969. i-iv + 14 p., tables. UNCLASSIFIED

A method for direct numerical evaluation of the cumulative probability distribution function from the characteristic function in terms of a single integral is presented No moment evaluations or series expansions are required Intermediate evaluation of the probability density function is circumvented. The method takes on a special form when the random variables are discrete.

1. Statistical Functions
I. Nuttall, A.H.
II. Title
III. ZF 11 112 002